INTERGEL® Adhesion Prevention Solution

STATISTICAL EXPERTS CONSENSUS REPORT

August 6, 2001

Theodore Colton, Sc.D.
Professor and Chair Emeritus
Department of Epidemiology and Biostatistics
School of Public Health
Boston University

Steven Piantadosi, M.D., Ph.D.
Director of Biostatistics
Professor of Oncology, Epidemiology and Biostatistics
Johns Hopkins University School of Medicine

Donald B. Rubin, Ph.D.
Professor and Chairman
Department of Statistics
Harvard University

INTRODUCTION

The present panel of statistical experts was convened by the Sponsor (Lifecore Biomedical, Inc.) to evaluate the design and data analysis of the pivotal trial for INTERGEL® Adhesion Prevention Solution (INTERGEL®), prompted by questions posed to the Sponsor by the Office of Device Evaluation (ODE) and written positions taken by ODE staff. The experts independently reviewed the submission that is the subject of consideration by the Food and Drug Administration, the Premarket Approval Application (PMA) as amended (June 2, 2000), along with positions taken by ODE in correspondence with the Sponsor and in internal memoranda with regard to the trial design and data analysis issues. Specifically, we relied upon the not-approvable letter issued by ODE on November 15, 2000 and the Statistical Review of Lifecore's Clinical Study prepared by ODE (provided by ODE to the Sponsor in May 2001) as the source of relevant information regarding the statistical review by ODE of the PMA as amended.

This report provides our consensus opinion, namely, that this clinical trial provides valid scientific evidence upon which to base conclusions regarding the effectiveness and safety of INTERGEL® prior to marketing. We concur with the Sponsor that the trial is

well-designed and that the analysis of the study as described in the PMA as amended is scientifically sound. This analysis proceeded as described in the study protocol; additional analyses were carried out by the Sponsor at the request of ODE and in accord with appropriate statistical and scientific practices. We have independently assessed the validity of the trial conclusions and find that these results are statistically robust and are sufficient to support approval. We did not identify any methodological or statistical issues of concern that are sufficiently critical or problematic to undermine the validity of the clinical trial results as presented by the Sponsor in the PMA as amended.

CONSIDERATION OF THE STATISTICAL ISSUES

We evaluated the entire methodology and data analysis of the INTERGEL® clinical trial, including questions posed by ODE on the statistical issues. The issues of particular interest to ODE were: adherence to the data analysis plan in the study protocol; the potential bias due to incomplete ascertainment of data at second-look surgery; the potential for bias due to consideration of the "evaluable" population vs. the "intention-to-treat" population; the combining of the data from the U.S. and Europe; the statistical power of the study; and, the calculation of the primary endpoint (the AFS score) from the adhesion score at all anatomical sites (mAFS score). Our assessment, analysis, and conclusions are provided below.

(1) DESIGN OF THE INTERGEL® PIVOTAL TRIAL

The pivotal trial of INTERGEL® was designed in 1994/1995 and completed in 1999. The study incorporated critical design features to eliminate bias in surgical clinical trials. These elements of a sound trial design include: a multi-center study; strict randomization at each investigational center; and, treatment masking (blinding) of the surgeon, evaluator, and patient. The trial required that each patient undergo two invasive procedures, a laparotomy and a subsequent laparoscopy ("second-look"), during which adhesions were directly assessed using a standardized assessment instrument. Subject accrual to this trial was slow due to the inclusion/exclusion criteria imposed by the protocol and ethical constraints. This required the use of 16 study centers (11 in the U.S. and 5 in Europe). All investigational centers followed the same study protocol and procedures.

The study was designed to test the hypothesis that, compared to lactated Ringer's solution (control), INTERGEL®, as an adjunct to good surgical technique, reduces the risk of post-surgical adhesions. The study, therefore, was an evaluation of the incremental impact of an adjunctive surgical treatment using a double-blind randomized experiment. This is a particularly challenging effect to measure with sufficient precision in trials necessarily limited in size due to logistic and/or ethical constraints.

Our opinion is that this trial clearly meets applicable scientific and regulatory standards. The results of the INTERGEL® pivotal trial are supported both by an earlier pilot study,

and research in animal models. Altogether, these data demonstrate consistency and biological plausibility --two important features that enhance statistical findings and conclusions regarding the safety and effectiveness of medical devices that are based on a single pivotal trial.

(2) ADHERENCE TO THE STUDY PROTOCOL

The PMA as amended presents all of the analyses planned in the study protocol (see clinical study report in PMA as amended, Appendix D). The statistical analyses performed by the Sponsor to assess whether this product is effective and safe are sound. All secondary endpoints were prospectively defined in the protocol and these analyses of secondary endpoints were provided to ODE in the clinical study report. We do not find that the multiple endpoints assessed and evaluated in this trial present any unique statistical issues of multiplicity as has been suggested by ODE. The analysis provided by the Sponsor in the PMA as amended, with regard to the consideration of multiple measures of effectiveness, is appropriate.

The PMA as amended includes an analysis of the AFS score, which is a measure of adnexal adhesion incidence, severity, and extent. This measure comprises the data gathered prospectively at 10 of the 24 anatomical sites evaluated for all patients. These additional analyses, although performed after the clinical trial was complete, do not represent inappropriate post-hoc analyses, retrospective analyses, or "data-dredging." The PMA as amended described a new indication for use, but not a new dataset. The data are not retrospective. Furthermore, the additional analyses were performed at the specific request of ODE (in a major deficiency letter to the Sponsor dated December 7, 1999) in accordance with standard practices and based on a sound clinical rationale in accord with the study hypothesis as articulated in the PMA as amended (Meinert 1986, p. D287).

(3) STATISTICAL POWER

Statistical power is a probability statement about a hypothetical treatment effect. In the INTERGEL® pivotal trial, the observation of statistically significant differences between treatment and control for any endpoint is not affected by the original power calculation in the study protocol.

The sample size for this pivotal trial was determined based on differences in the mAFS outcome observed in a pilot study. (The mAFS score is the AFS scoring method for adhesions applied to all 24 sites evaluated in each patient). The original power calculation was based on standard methods and reasonable assumptions from the pilot study. Had there been a more precise estimate of the true standard deviation of the primary outcome for the pivotal trial, one could have planned a smaller trial.

When the trial was complete, the difference in mAFS scores seen in the study (evaluable population) was smaller (1.0 unit change) than what had been anticipated when the study was designed based on the pilot study results (2.1 units change). However, the standard deviation was also smaller (1.5 for INTERGEL® and 2.2 for control) compared to the standard deviation (5.0) upon which the sample size was estimated, which led to greater precision of estimation than had been anticipated. There is no rationale for questioning the validity of a trial because the treatment effect and/or variance are smaller than that hypothesized before the study was undertaken. This trial assessed the incremental benefit provided by an adjunctive treatment and found it to be statistically significant. It may well be preferable to have a more precise measure of the outcome of interest with less variation, than a large effect with more variation. We leave comments on the medical importance of the statistically significant results to the judgment of clinical experts in the field.

(4) COMBINING THE U.S. AND EUROPEAN SUBJECTS ("Pooling")

The Sponsor provides a sound rationale for the use of data from all clinical trial sites, with which we generally concur (see PMA as amended, Section III. 3.0 Justification for Use of Data from All Trial Sites). Patients were enrolled or excluded from participation under the same inclusion/exclusion criteria. Furthermore, identical trial procedures in the U.S. and Europe were employed to implement the trial. A critical factor supporting the combination of data across continents and/or centers is that randomization was stratified by center. Variation in baseline characteristics by center is expected and is among the reasons for doing a multi-center trial, and for randomizing patients separately at each study center. Although in this study, differences in characteristics were observed between continents for some baseline variables (as one would expect), there were no differences in baseline characteristics between treatment groups, which is critical. The strongest justification needed for combining data is stratified randomization, which produces an unbiased estimate of treatment effect from each center.

When the data were stratified by both continent and adhesiolysis category (both variables of concern identified by ODE) the Breslow-Day test of Homogeneity was not statistically significant, indicating that the continent-specific estimates and the adhesiolysis sub-group estimates were not heterogeneous. To control for the effects of continent and adhesiolysis category, the patients were stratified by the variables of concern: continent and adhesiolysis category. When the primary analysis was stratified by these variables, the differences between treatment and control were statistically significant.

(5) INCOMPLETE ASCERTAINMENT

The INTERGEL® pivotal trial consisted of 303 patients entered, with 281 treated (143 INTERGEL® and 138 control). Of the 281 patients treated, 265 completed the study

(131 INTERGEL® and 134 control). One patient was lost to follow-up (INTERGEL® group). The remaining 15 patients were duly accounted for according to the study protocol (which specified that the "principal investigator will provide an observation of the patient's condition, and whether any complications were noted"), although one of these women refused to complete her medication diaries. Fifteen patients did not have the second-look laparoscopy due to pregnancy (1 patient), physician decision (1 patient), or the patient's decision (13 patients). There were no discontinuations due to a reported adverse event.

The pivotal trial results reported by the Sponsor in the PMA as amended relies on the evaluable study population (265 subjects). The statistically significant results, in favor of INTERGEL® over control, for the evaluable population, were supported by four separate imputation analyses to assess the validity of the conclusions (see PMA as amended, Section III. 4.0 Analysis of Incomplete Ascertainment Subject Data). These analyses were: Imputation by Control Group Failure Rate; Imputation by Informed Censoring; Imputation by Worst Case AFS Score; and, Additional Imputations Using Various AFS Scores. The approaches applied by the Sponsor are in accord with standard practices and included appropriate clinical rationale (per the International Conference on Harmonization Guidelines for Statistical Principles of Clinical Trials. ICH, 1998). The results based on these imputation analyses support the validity of the results obtained from analyses of the evaluable population. Of note, these sensitivity analyses did not "ignore those subjects lost to follow-up"; on the contrary, the four imputations developed were specifically designed to consider the impact of these subjects without second look data on the evaluable population results.

(6) INTENTION-TO-TREAT ANALYSIS

The FDA required the Sponsor in the study protocol to include a different type of imputation analysis to account for missing data other than the four separate analyses provided in the PMA as amended and cited above. This analysis (which was termed incorrectly by ODE as the "intent-to-treat" analysis), required that all patients who did not have a second-look laparoscopy be considered as failures and assigned the worst possible adhesion score. When this was done with the primary endpoint, as specified in the study protocol for the original intended use (the mAFS score), the results were statistically significant.

This analysis, based on worst-possible value imputation, although statistically significant, is neither mainstream nor scientifically defensible because it is overly conservative. There is substantial information in the dataset regarding patients who did, and did not, have a second-look laparoscopy. Only one patient was truly lost to follow-up, and a second refused to provide complete information. The remainder of patients who refused the second-look laparoscopy did so because they did not want to be subjected to a second surgery, were feeling well, or were pregnant. This information, along with other clinical variables, should be considered in any scientifically sound imputation.

We recognize that the failure to consider the intent-to-treat population may obscure important findings or lead to incorrect conclusions. Nevertheless, we find that analysis of the evaluable population in the INTERGEL® pivotal trial as provided by the Sponsor is appropriate. The lack of data on the small number of subjects who failed to have second-look laparoscopies did not materially alter the analysis of the results of the trial.

Finally, we note that the Sponsor did adhere to the data analysis plan in the study protocol, which specified three different populations for consideration: all patients treated (the intent-to-treat population); all patients for whom a second-look laparoscopy was conducted (the evaluable population); and all patients treated excluding those who refused a second-look laparoscopy. The protocol did not specify, nor in our interpretation imply, that the "primary" data analysis was to be carried out on the intent-to-treat population. To the contrary, the protocol specifically anticipated the refusal of study subjects to undergo a second invasive procedure by specifying a population that excluded these subjects.

Although we found the evaluable population to be an appropriate basis for analysis of the results of the INTERGEL® pivotal trial, for the reasons cited above and as established by four separate and appropriate sensitivity analyses, we independently undertook an analysis of the entire randomized population (a true "intention-to-treat" analysis) in order to address this statistical concern posed by ODE. We note that ODE provided an "intent-to-treat" analysis of the PMA as amended to the Sponsor in May 2001 (Statistical Review of Lifecore's Clinical Study, Table 3), but we do not feel this is an appropriate methodology.

The multiple imputation methodology we utilized for an analysis of the intention-to-treat population has been validated in numerous publications (see for example, Rubin 1996; Rubin and Schenker, 1991). A complete description of the methods and results are available upon request. Briefly, the imputation method was applied in a completely blinded manner to avoid any possibility of bias entering the imputations. Because the clinical trial had been completed, the specification of the technique did not involve any knowledge of which group was the active treatment group and which was the control. That is, the imputers and the designers of the imputation method were blinded to treatment labels. Finally, because of the critical nature of the imputations and the need for complete objectivity, it was decided that the imputation method to be used should not have available any second-look data whatsoever. This decision was implemented by using a nonparametric matching method for imputation. Subjects who had observed second-look variables were selected based on pre-specified clinically relevant baseline variables and matched to those women without observed second-look based on these baseline variables. The donors then donated their observed second-look values to those women without observed second look data. Based on the imputed databases, pre-specified analyses were run independently by the original Lifecore statistical consultant and the independent company that performed the blinded imputations.

The results of this scientifically sound and rigorous intention-to-treat analysis strongly confirm the findings submitted by the Sponsor in the PMA as amended. Women who presented in this trial with no, minimal or mild adhesions who were treated with INTERGEL® were more likely to remain in that category than women treated with lactated Ringer's solution. Moreover, women with moderate or severe adhesions at baseline, when treated with INTERGEL®, had--at worst--mild adhesions at second-look, whereas those treated with lactated Ringer's solution were more likely to present with moderate or severe adhesions. The conclusion, based on the intention-to-treat population following scientific imputations, is that INTERGEL® is superior to lactated Ringer's solution for all subjects, no matter what their baseline adhesion score (Relative Risk >5; p=.0003). These results are compelling for their consistency, degree of significance, and small degree of variability or uncertainty. These conclusions are consistent with the findings reported for the evaluable population. The findings are not dependent upon continent or baseline adhesions scores.

(7) CALCULATION OF AFS SCORE

We reviewed the clinical expert report on INTERGEL® safety and effectiveness (DeCherney et al, 2001) with respect to the calculation of the AFS score and find no unaddressed statistical issues of concern. As stated in the clinical report, sensitivity analyses were conducted to test the potential variability of AFS scores computed in the trial. Even if the method of calculating AFS scores in the INTERGEL® trial introduced additional variability, this could only increase the likelihood of non-significant results and could not affect the existence of significant findings. Variability introduced by score computation is not a concern with statistically significant results, as were obtained in this trial.

CONCLUSIONS

We conclude that the INTERGEL® pivotal trial results as presented by the Sponsor in the PMA as amended are appropriate and reliable. The study demonstrates a statistically significant difference in AFS scores compared to control with a high degree of confidence (a 5-fold lower risk of moderate/severe adnexal adhesions), which are supported by statistically significant analyses of secondary outcomes. These results are also consistent with those in the pilot study and research in animal models. The results reported by the Sponsor in the PMA as amended for the evaluable population are well-supported by four separate imputation analyses. We have independently verified these conclusions with an appropriately designed and conducted intention-to-treat analysis based on scientifically imputed second-look data. The data analyses presented by the Sponsor in the PMA as amended are, in our opinion, robust and certainly adequate to support conclusions regarding clinical significance by appropriately qualified clinical experts.

REFERENCES

DeCherney, A.H., Ellis, H., Faro, S., Holmdahl, L., Malinak, L.R., Martens, M.G., Mastroianni, Jr., L., Sever, J.L. 2001. INTERGEL® Adhesion Prevention Solution. Blue Ribbon Panel Consensus Report on Safety and Effectiveness.

International Conference on Harmonisation (ICH). 1998. Guidance on Statistical Principles for Clinical Trials; Availability. Federal Register (63) 49583-49598.

Meinert, C.L. 1986. Clinical Trials. Design, Conduct, and Analysis. New York: Oxford University Press. Pp. 214-215.

Rubin, D.B. 1996. Multiple imputation after 18+ years. J. Am. Stat. Assoc. 91(434):473-489. With discussion, pp. 507-515, and rejoinder, pp. 515-517.

Rubin, D.B and Schenker, N. 1991. Multiple imputation in health-care databases: An overview and some applications. Statistics in Medicine 10:585-598.

EXPERT CONSULTANTS

Theodore Colton, Sc.D.
Professor and Chair Emeritus
Department of Epidemiology and Biostatistics
School of Public Health
Boston University

Steven Piantadosi, M.D., Ph.D.
Director of Biostatistics
Professor of Oncology, Epidemiology and Biostatistics
Johns Hopkins University School of Medicine

Donald B. Rubin, Ph.D.
Professor and Chairman
Department of Statistics
Harvard University